PERCEPTUAL CARTOONIFICATION IN MULTI-SPATIAL SOUND SYSTEMS

Peter Lennox

University of Derby
Creative Technologies Research Group
Derby DE22 3AN, United Kingdom
p.lennox@derby.ac.uk

Tony Myatt

University of York
Music Research Centre
York YO10 5DD
Am12@york.ac.uk

ABSTRACT

This paper describes large scale implementations of spatial audio systems which focus on the presentation of simplified spatial cues that appeal to auditory spatial perception. It reports a series of successful implementations of nested and multiple spatial audio fields to provide listeners with opportunities to explore complex sound fields, to receives cues pertaining to source behaviors within complex audio environments. This included systems designed as public sculptures capable of presenting engaging sound fields for ambulant listeners. The paper also considers questions of sound field perception and reception in relation to audio object scaling according to the dimensions of a sound reproduction system and proposes that a series of multiple, coordinated sound fields may provide better solutions to large auditorial surround sound than traditional reproduction fields which surround the audience. Particular attention is paid to the experiences since 2008 with the multi-spatial The Morning Line sound system, which has been exhibited as a public sculpture in a number of European cities.

1. INTRODUCTION

What can make artificial spatial sound more believable, informative and interesting? We approach this question by recourse to a classic formulation of the problem of what perception *is*, and is *of*: Plato's Cave Allegory [1]

Current "artificial" sound environments are far more detailed than the simple shadows and echoes described by Plato, yet seem less convincing because the presentations are not linked comprehensively to a causal foundation.

Complex artificial audio-spatial environments are in their infancy and there is a need to develop ontological, perceptual and aesthetic foundations for this work.

In this paper, we begin with two hypothetical tenets:

- That what we perceive is *true*, but not the *whole* truth: "Plato's cartoons".
- That perception is modular, not monolithic.

We then hypothesise that the particular "cartoons" we observe in real environments stem from perceptual modules evolved to select real causal distinctions. This is extended, speculatively, to the notion that we can explore the nature of these perceptual modules through interaction with artificial "cartoons". However, in this case it is important to avoid undue constraints by assuming that laboratory findings (which are necessarily constrained artificial environments) defines what we *can* know about auditory spatial abilities.

We have constructed a series of novel artificial spatial sound displays to explore the possibilities for spatialisation in unconventional listening circumstances. Particularly, we are interested in scale and control of the large-scale and detailed aspects of the listening environments.

In artificial spatial sound, as in any artificial environment, the guiding principle is that physical properties are manipulated to evince experiential qualities.

We focus on the latter, using the principle of the "perceptual cartoon" which we define as:

> A sparse set of relationships, represented in cognition, which couples directly with essential causal features in a perceiver's environment.

In such a context, "spatially accurate audio" should be measured in terms of perceptual rather than physical accuracy. For spatial audio, the term "spatial" should be closely defined in terms of the target knowledge structures.

2. PLATO'S CARTOONS

In the Allegory of the Cave, Plato discusses questions of veridicality in perception. The human subjects of his thought experiment are artificially constrained, such that their only perceptual experiences are of shadows which are deliberately cast on a cave wall, combined with ingeniously coincident reflected sound. The subjects cannot move, turn their heads or see each other.

Plato asks whether, if, for the first time in his life, a captive is then allowed to turn round and see the actual objects that have cast those shadows, and is then escorted out of the cave to see the real world, they will be able to use perceptual abilities developed in the cave.

That which is available to perception directly is a small fraction of the world. This is now uncontroversial. In the allegory, what has been shown is not that what is available through sense presentations is untrue, but simply that it is incomplete. The limited shadows are still causally connected to the 'real' state of affairs. Plato's shadows are not fanciful artefacts but "cartoons" of reality. They are constituted of some of the essential properties of the underlying reality. "Essential" here means: "at the level of description appropriate to our perception". With respect to Plato, the underlying causal context is divided into comprehensible items by perception. Items such as places, features, objects, organisms, events and relationships, even opportunities and threats, are "cartoons" that seem entirely real to us simply because they are real…to *us* as human beings.

Plato asks whether the released subject may, in time, come to understand this new level of reality perceptually.

He does not raise the distinction between ontogenetic and phylogenetic development, but clearly, many of our abilities require both. Language, for example, requires certain capacities to have developed at species level but also though exposure to language use. We also continue to extend perception through mediating technologies which allows us indirect access to levels of description that are too large, too small, too fast or slow, and beyond our direct sensory capabilities.

2.1. Ontology in artificial environments.

Perception in an artificial environment is broadly defined as a differentiated sensory array where information transactions are managed partly or wholly by design. Users' experiences in such environments rely on the designers implementing some theories (implicit or explicit) of perception. These may be sensory theories elicited in previous psychophysical experiments (perhaps even with their experimental subjects constrained as stringently as in Plato's cave in some instances) that outline constraints on dynamic, frequency and temporal ranges and acuities.

There remains, however, the ontological tangle of what it is that is being presented. Focusing on patterns that interact with organs of sensation is the wrong level of description to capture the notion of complex information being conveyed. Plato's shadows are real because they are causally connected to some real state of affairs.

In some cases, signals in an artificial display are strongly linked to something real. For example, the audible display of a Geiger counter emits simple clicks whose rate of occurrence varies with the intensity of the ionising radiation present. Although the clicks are not *like* anything, their rate is causally linked to radiation intensity level. The information is meaningful, succinct and intuitively easy to use, providing a sophisticated extension to perception via a simple technological coupling between an imperceptible field and a perceptible one.

In written language (excluding pictorial languages), patterns presented to sensation are not necessarily *like* the items, scenes and stories depicted, yet complex information-transactions are feasible.

3. CARTOONS, PHYSICAL, ARTIFICIAL AND PERCEPTUAL

We have argued that it is implausible to suggest that a given perceiver is in contact with *all* the information available at a given position-and-time. Therefore, perception logically entails information reduction, implemented at structural levels as well as processing levels. What results are perceptual cartoons, at a certain scale or resolution of the ongoing causal context. These interact in the definition of perceptual context systems to produce a holistic grasp of the environment. Perception must achieve salient analysis of complex environments in a timely fashion.

Hence, perceptual cartoonification can be distinguished from the sort of artificial physical cartoonification that Gaver [2][3] proposes, in that it provides a plausible explanation of why physical cartoons can work at all. The claim here is stronger: cartoons *are* what we perceive. The physical properties of any event, object or entity are innumerable; even in the position that perception is *direct* (as in the *ecological approach* [4]}, it cannot be complete.

For instance, the auditory experience of a fast-moving item cannot consist of the apperception of an item that occupies successive locations; what is available directly is motion, direction and (possibly) size.

The concomitant assertion concerns perceptual cartoons' appropriateness, selected for by phylogenesis (primarily) and ontogenesis (secondarily). Perception has evolved to pick up essential physical features. Hence, artificial attempts at physical cartoons should match in terms of *sparseseness*.

3.1. Artificial Cartoons

There are countless examples of artificial physical cartoons that appeal effortlessly to *intuitive accessibility* [5]; their nature can illuminate the underlying perceptual processes at work. An obvious example is an outline drawing of an animal. The animate creature will never actually look exactly like the frozen representation; the shape is an idealisation.

Outline drawings can even capture something about movement, as though movement consists of a sequence of frozen-moment static shapes. In some examples, 'background' surrounding scenery is depicted as blur whilst the foreground, moving item is depicted in high resolution; this mimics visual focus.

Motion can be cartoonified without outline shape. The Biomotion Lab in Canada and Germany (see [6]) uses motion capture technology to track the relative displacement of 15 key points on human walkers. Observers can then (surprisingly easily) describe the sex, age, weight and mood of the walker from the behaviour of 15 moving dots on a graphic display. These attributes can be represented and manipulated algorithmically, [7][8] to adjust viewers' perceptions. Note that, before the display is animated, the 15 dots mean little and the perception of a human does not arise.

3.2. Perceptual Cartoons

Cartoons are not necessarily sense-mode specific, concepts of “now” for grasping the flow of events, and “what-and-where” for understanding *objects* in *locations* are examples. Indeed, the fundamental notion of an *object* (divorced from particular behaviour) is a primitive cartoon.

There is evidence supporting notions of quasi-discrete auditory ‘what and where’ neural processing streams [9], lending weight to the idea that perception cartoonifies sense-data in terms of *what* things are and *where* things are. We use an initial coarse classification of *thing* (organism or object), *place* (the shared environment in which perceiver and things perceived have location), *feature* (that might afford access, climbing, hiding, hefting or obstacle), *relationship* (nearness, direction, hidden behind, on top of, in front of… a feature), *dynamic behaviour* (coming, fast, passing, departing, bouncing scraping, accelerating) and finally, *intentionality*.

Cognitive spatial frames of reference [10] are cartoons of spatial layout with respect to potential interaction. They tend to be centred on the perceiver or part thereof (e.g. head-centric) or some anchoring feature in the environment. They can facilitate interaction without necessary recourse to recalculating each and every spatial relationship in the environment.

“Attention” is cartoonification; information reduction is achieved by concentrating on particular causal contexts at the expense of others, exemplified in phenomena such as *change blindness* (reviewed in [11]), where large changes in the visual scene can go undetected, and *inattentional blindness* [12][13][14] where significant items fail to enter conscious representation (even if in the centre of the visual field). Studies of drivers using mobile telephones highlight depleted awareness of items in peripheral vision (e.g. [15])

Apparently, attention-as-cartoonification of the causal scene is an intrinsic feature of perception and causal features compete in terms of perceptual significance.

We proceed from the premise that perception actually cartoonifies the causality in the environment local to the perceiver, representing the *meanings* that are relevant to the percipient’s welfare. Spatial perception is thereby cast as a subset of *place-perception*. In this view, rather than auditory perception proceeding from an accurate (as possible) three-dimensional representation replete with precisely localised sources, it is subjectively distorted toward representing the immediate future as it might impinge on the perceiver.

Cartoons perform the function of reducing cognitive workload, liberating resources for high-significance tasks. So, *recognizing* a particular object, item or feature allows cognition to reduce the resources devoted to analysing sensory input from that item. For example, once one has recognized a *dog*, one can access the range of behaviours of which that class of item is capable and treat it accordingly. Whilst the actual dog conforms to the behaviour anticipated in the cognitive model of it, sensory processing is minimized. Should anomalous behaviour be detected, the dog will return to the forefront of attention.

We focus here on the sense-specific features that are available for caricature. Although the underlying causal features (organisms, objects, places, boundaries, obstacles, throughways) are not sense-mode specific, they do afford mode-specific features. In vision, ubiquitous ambient light reflects off material features fairly uniformly (in terms of brightness); what is available at the retina is a simultaneous topographic map of everything that is in line-of-sight now. The challenge for visual cognition is to parse this into meaningful entities’ boundaries. To even recognise an important potential predator or competitor, it must be identified as a thing-that-may-be-capable of certain behaviours. Surface texture, colour, outline, edges, shape and size are visually available. Proximity (to the perceiver) is available to vision via texture gradients, perspective and occlusion. Estimates of future proximity can come from this in conjunction with estimates of the object’s motion-capabilities (how fast can those things move? Which way is it facing to start with?). Location (with respect to the percipient and other environment-features: is there a route for escape?) is apprehensible. Very fine properties like facial expression and body language can also be elicited. We do not expect exact analogues in audition. The first problem that vision must solve (dividing the world into thing and not-thing) is, in audition, fairly unchallenging; sound-sources are things and reflected sound is not-thing (i.e. an environmental characteristics). Hence, surface properties are less relevant, as are image size and precise direction. Internal properties (hollowness, rigidity/softness, heaviness, are more apprehensible in audition. Behaviours (vocalisations, locomotion, acceleration etc) are excellently apprehensible without line-of-sight.

4. SPATIAL AUDIO APPLICATIONS

In audition, “place” is the causal backdrop in which things (sound sources) and the perceiver find themselves. The task of perception is to understand where things *will* be, soon; prediction. Thus, information reduction through cartoonification according to perceptual significance should extract features likely to become important for prediction. Reassuringly distant things should occupy lower positions in an attention-hierarchy than uncomfortably close things. The attributes of the latter should be especially noticeable. Things coming toward me (the perceiver) should capture attention more than things departing, things that sound as though facing toward me, likewise. Things that display organism-like characteristics (vocal utterances, sudden acceleration, and change of direction) would occupy the perceptual foreground.

Place-characteristics, by contrast, sit easily in the perceptual background. These would include size and shape, enclosed, partially enclosed or open, cluttered and busy or empty and calm. Such properties must be extracted from the conflated signals arriving at the eardrums specifically as subjects for *inattention*. It is known that perceptual performance for localising sources improves as the reverberant characteristics of the particular environment are learnt [16].

It is an important feature of our approach that background place characteristics must be cartoonified well *enough* to support plausible localisation of things. Location is the property of places rather than of things. We have heard many synthetic sound fields where location is treated as a property of things, so that a simple directional vector is depicted for each source, and a collection of sources with different vectors amounts to 'surround sound'. Perceptually and aesthetically, this falls short of what could be achieved in artificial spatial sound fields.

4.1. Extant Examples of Audio Cartoonification, Physical and Perceptual

In audio practice, using discrete microphones to capture sources and ambience exemplifies physical cartoonification. For example, sometimes, real sources are captured and ambience generated via convolution using impulse-response data taken from real places (e.g. [17][18]). In recent developments, recordings can be partially de-constituted into source-signals, their directional vectors, and reverberant material (e.g. [19][20][21])

The technique of deploying spaced microphone pairs to capture *spaciousness* in audio ambience comes at the expense of image stability (images of moving items can 'jump' from one speaker to the other with no smooth transition). By contrast, coincident Blumlein techniques, although giving better image stability, capture less of the sense of spaciousness. The tentative conclusion is that, to human physiology and cognition, thing-characteristics and place-characteristics are quite distinct – not only subjectively, but also in real physical terms.

Current examples of physical cartoonification include the uses of artificial reverberation to simplistically model place characteristics, manipulations of inter-aural differences (via pan-pot controls) to steer perceptions of phantom sources and synthesising perceived range (source-perceiver distance) by controlling the 'dry-reverb ratio'.

However, object size, orientation, types of movement (such as coming, passing and going) as well as interactions between synthetic objects and synthetic environment are all currently impoverished in comparison to their counterparts in real environments, except where sound fields are represented in the recording systems, e.g. with Ambisonic sound field recording techniques.

Note that things' virtues such as *position*, *separateness* and *size* are simply not applicable to place characteristics. They are properties of items that one *could* touch or circumscribe, the solid items in the world that stereo (Greek for solid) is supposed to depict. By contrast, although a place has size, it is of the sort that the percipient is *in*, surrounded by – it is not an image; there should be no point-source character to reflected sound.

The exception to this general rule is in the case of very large things such as the sea, a storm, the rain, an audience – all have an elusive, 'everywhere-and-nowhere-specific' quality in many reproductions. Some elements of place have a stable, locatable nature – these are items like an occluding wall, a doorway, a reflective wall; these equate to our notion of causal mid-ground features

To sum up, *thing*- but not *place*- character should seem tangible. Unfortunately, phantom images are elusive and lose their tangibility as soon as a perceiver moves within a local sound field, whilst the surround (ambience reproduction) loudspeakers can be too source-like.

4.2. Extending Audio Cartoonification

Whilst traditional approaches to spatial sound can suffice for a single well-placed listener (listening to the stereo or playing a computer game, for instance) the systems so designed don't have a suitable pedigree for large-scale listening circumstances. In dance clubs, stadia or concert halls, most listeners are not in the correct place and so the spatial sound cartoons are poorly received.

We have experimented with several alternatives approaches.

4.3. Larger-Scale Surround-Sound Systems

An interesting line of investigation stems from the observation that, with a qualified exception of Wavefield Synthesis, existing surround sound technologies do not 'scale up' to large listening areas without 'scaling up' the size of the reproduced sounds.

In collaboration with the event staging company Funktion One Ltd. at the Glade Festival [22], Lennox used an outdoor six-speaker horizontal-only surround array to explore the suitability of Ambisonic surround techniques for larger arrays. The powerful (40KW) system had a radius of approximately 25 metres. Some material was encoded through VST 2nd order ambisonic plug-ins (running in the VST host "Audiomulch"). Other material consisted of 1st-order natural recordings captured through the Soundfield microphones. Some of this material was 'close miked' – that is, the source-mike distance was considerably smaller then the speaker radius. Other materials such as aircraft flyovers were obviously more distant. The distant material, the synthetically encoded material and recordings made in very large buildings (cathedrals) were accurately reproduced. The close-miked material was interesting, but anomalous in terms of audio image size.

If one simply increases the radius of the loudspeaker array (and because, the larger an indoor listening area is, it is less likely to be anechoic), images simply stay at the circumference. Using a sound system with sufficient power to address the whole listening volume can (possibly because of a perceptual constancy for the speaker distance) produce perceptions of images that are larger than their true size. This was informally observed by many listeners, who reported Geese the size of dinosaurs [23], motorbikes 15m high and a flushing toilet cistern fit for a giant. Whilst on occasions this might be a useful property of such systems, the lack of control of the perception of image size is potentially problematic, indicating that some information is lost during the up-scaling to large surround arrays.

4.4. Periphonic Fields

In addition to the distortions of spatial intent occasioned by precedence effects for off-centre listeners (the majority in a large environment) where imagery – and even ambience – can be perceptually ‘pulled’ toward the nearest loudspeaker, there are the problems associated with masking-by-occlusion caused by other people’s bodies.

To address these issues, Lennox assembled a periphonic 2nd order ambisonic array, using 32 speakers on the surface of a geodesic dome structure; this was for an experimental music festival where a traditional soundstage was not required. Synthesised 2nd order material was supplied by composer Jan Jacob Hoffman [24] (whose background is in architecture) which depicted imaginary structures and places, some under considerable (and audible) tension. For this material, the height element was vital.

Hence, it was feasible to depict spatial material above the listener. Because of the logistics of mounting heavy low-frequency drivers (and the fact that occlusion primarily affects frequencies over about 500HZ), an additional 8 speaker 1st order ambisonic horizontal-only floor-mounted array was utilized for <110 HZ. Whilst, for a static perceiver, these frequencies offer little coherent information to directional hearing mechanisms, the amplitude gradients (at LF) available in an array of this sort were found to be quite perceptible for off-centre and ambulant listeners. Informal observations were that control of variations in perceptions of spaciousness was excellent. This is in line with Marten’s assertion that “…two subs are better than one” [25]. In this application, it would seem that eight are better than two, not least because the room, a typical large enclosed and reflective ‘shoebox’ was notoriously difficult, acoustically. The arrangement used tamed the room utterly. Since elevation discrimination is largely a high frequency affair facilitated by pinnae effects, the vertical discontinuity between frequency bands above and below 110 Hz appeared imperceptible. A further advantage was that crossovers were digital, inserted into the signal chain only in the horizontal 1st order components – ‘W’,’X’ and ‘Y’, followed by a decoder directionally aligned to the main system.

Various spatial formats (stereo, 5.1, octophonic, 1st order ambisonic, 2nd order ambisonic) were converted to either 1st or 2nd order ambisonic (using 1st or 2nd order pan controls), subsequently decoded to the dome.

Notably, Hoffman’s material, encoded using Csound, featured finely crafted haloes of early reflections around source elements. These served to reinforce the “real source in an *actual* environment” impression of what was, after all, entirely synthetic. The perception of movement of sources seemed to be excellent. This provides a good example of the principle of managing the ‘thing-place’ relationships

Overall, the system served the listeners well, especially those positioned at the circumference of listening area – normally a zone where audio spatial illusions are quite unsatisfactory. As an aside, the sound stage sounded excellently spatial (though less precisely informative) from positions outside the array; a similar observation has been made by Malham and Myatt [26]

4.5. Concentric Fields

Lennox also constructed a hybrid system comprising an inner (10metre diameter) 4 speaker array: 1st order ambisonic, horizontal 360 degree spatiality, and an outer (30 metre) 12 speaker array: horizontal 1st order ambisonic. Some material was B-format encoded material either captured by a Soundfield™ microphone or synthetically encoded from multi-track material. Other material was conventional stereo or mono material, encoded through analogue encoders. Additionally, parallel signal chains allowed discrete access to each speaker via sub-group sends from the sound desks.

By convention, the inner array was generally used for dry source material whilst delayed and reverberated versions of the same material were fed to the outer array. This arrangement explored the relationships between the central stage (where ‘things’ were) and the distant region (for ‘not-things’ like storms, huge hall sounds, helicopter-over-the horizon, etc). The arrangement also permitted cartoonification of approaching and departing elements (i.e. coming here, leaving here, fading into the distance)

A more sophisticated version of concentric fields was used in a live performance work, *untitled 3*, at the Staatliche Hochschule für Gestaltung Karlsruhe, 22nd June 2007. This system used both vector-based amplitude panning (VBAP) and 3rd order Ambisonic panning to control the 43 speaker system. The outer 2 arrays also featured vertical heterogeneity to facilitate *falling* and *rising* items. The performance space was large, allowing the loudspeaker array to be divided into three systems: an inner circle, a mid-ground geodesic hemisphere configuration and an outer rectangular cube. This allowed a series of creative audio experiments relating to sound objects' interaction with a defined audio environment, along with experiments in establishing spatial audio context for sound movement, i.e. within a spatial audio environment and relative to other moving sound sources. The significant distances between nested field allowed listeners to explore the sound environment perceptually to some extend, since moving themselves or their head, revealed more information about the spatial location of sound sources on the near, middle and far distant reproduction systems (unlike a conventional surrounding sound system where moving elicits less spatial information).

4.6. Multi-spatial arrays

In 2008, Myatt developed the concentric array concept to include concentric and adjacent sound fields to constitute a *multi-spatial system system.* This was facilitated through a project by Thyssen-Bornemisza Art Contemporary, who are engaged in the development of a series of major art pavilions. They commissioned artist Matthew Richie, along with a team of architects and engineers, to create an anti-pavillion, The Morning Line. This was conceived to have a spatial sonic identity from its conception and to become a collaborative project with composers and sound artists, and now a venue for new multi-spatial compositions. Myatt realised the audio system for this project and led the team who developed all the audio authoring and controlling software, designed the loudspeaker configurations and the spatial encoding/decoding strategies.

Fig.1. The Morning Line, Eminönü Square, Istanbul (Photo: Tony Myatt)

The Morning Line uses a 47 loudspeaker system (41 Meyer MM-4XP (waterproof) + 6 channels LFE (12 x Meyer MM-10) . As illustrated in *Figure 1*, the structure has an open architecture, allowing sound to permeate throughout. It is designed for ambulant listeners. The speaker system is divided into 6 sound "zones" (*Fig. 2*) of full with-height reproduction on a near human scale (the sound zones surround a small number of local listeners and are approximately 5-8m in diameter). All adjacent zones can be heard from within any one area, but the local zone is most perceptually significant.

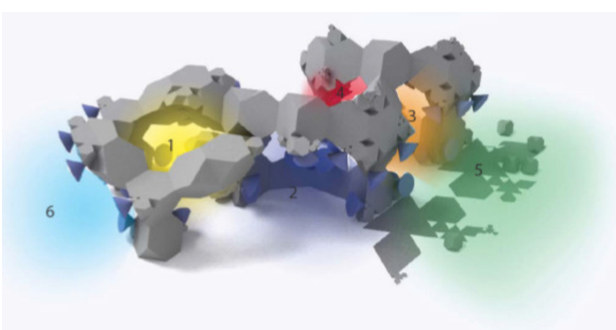

Figure 2. The Morning Line Rooms (Image: Aranda/Lasch)

The authoring software and system implemented a variety of sound encoding strategies. Decoding to loudspeakers is achieved largely using Ville Pulkki's VBAP [27] plus a variety of bespoke panning algorithms. Up to 100 sounds can be localised, placed or moved according to defined trajectories independently within one or many zones, or can be spatialised across the structure as a whole. Materials recorded in surround (e.g. multi-mic or Ambisonic Sound field) have also been used to establish sound fields in certain areas. A number of simultaneous audio spatialisation strategies are used.

Listeners are able to move and change their listening perspective throughout the structure.

During the initial installation of this work, many observers commented on their perception of the whole structure as having a clear sonic architecture, a sound mass perceived as a whole from a distance, constituted by multiple moving, oscillating or static sounds elements, outlining the structure of its abstract architecture and adjacent sound fields.

Equally, perception within any single zone places the listener within a local context of perceptually significant sounds or sound trajectories, but simultaneously within the audio environment of the whole structure, which defines the larger scale audio environment in three dimensions.

Over-here and *over-there*, close, present, *thing*-, *place*-, perceptually significant and insignificant materials can all be represented simultaneously across the loudspeaker arrays.

The ability to explore and to move through this sound environment seems key for many listener's apprehension of spatial audio characteristics.

The Morning Line was previewed at the Seville Biennial of Contemporary Art in 2008, exhibited by the Centre for Andalusian Contemporary Art in 2009, and shown in Eminönü Square, Istanbul, Turkey as part of the European City of Culture Festival 2010. It will be installed in Vienna from June 2011.

Approximately twenty-five composers have, to date, been commissioned by T-B A21 to create new works for this system and to explore this approach to multi-spatial sound presentation. Details of individual compositional approaches and materials are beyond the scope of this paper. Further documentation can be found at
http://www.tba21.org/pavilions/83?category=pavilions

4.7. Cellular Fields and Matters of Scale

The Morning Line addressed the problem of 1-to-1 spatial mapping (where the listening area is similar in size to the captured or synthesised spatial scene) by adopting a cellular/multi-spatial approach. We felt this approach promising because it could (theoretically) lead to an upwardly-scalable solution, where a large place could be depicted in an equally large place using a series of local sound reproduction fields (this also allows The Morning Line to be reconfigured, a feature also facilitated by its architectural design).

Lennox has also explored the possibility of navigability within such loudspeaker arrays, initially with a 3-cell approach, using three sound fields captured by three Soundfield ™ microphones, approximately 25m from one to the next. For initial simplicity, the microphones were arranged in a line rather than a triangle (so one horizontal dimension is truncated as for 'normal' surround sound). Material consisted of choristers and organ in a cathedral. The microphones represented the centre of the congregation area (though no congregation were present), the centre of the choir, the altar some 20m beyond the choir.

The recorded multiple sound field was subsequently reproduced in a considerably smaller venue but the centre of the cells (equivalent to the microphone position) were spaced by only a slightly scaled down 15m. Horizontal-only 1st order material was conveyed using 8 speakers per cell, whilst LF (<110Hz) was displayed by a single 4-speaker (1st order horizontal only) cell covering the whole of the listening area. More detailed display was eschewed due to resource constraints. With these constraints relaxed, attention would have been paid the height information which (on small-scale audition) was noted to be particular impressive.

On replay, the direct to indirect signal ratio at a given listener-position matched those for a listener in an equivalent position in the original environment

Further, precedence effects were properly appealed to. That is, the temporal relationships between 'most direct' and 'less direct' (a more appropriate term when discussing real environments, where some elements of the sound field – e.g. occluded objects – do not have a direct component for all frequencies) remain physically consistent, even during perceiver motion. Hence, the weak phantom image problem seems amenable to control. Likewise, *range* (apparent-source to perceiver distance) appeared consistent, to give an impression of approaching a sound source, in contrast to normal surround sound implementations.

In relation to perceptual impressions: it was particularly impressive to be able to 'walk through' the cathedral, passing through the congregation area, with the perception that one was approaching quite tangible sources. As one passed through the choir-cell, individual sources traversed through azimuth, front-side-back, and finally one could, as it were, leave the choir behind, moving onto the reverberant altar area.

However, as is often the case with first-order Ambisonic material, there was a slight vagueness of image definition. After the initial promise of near-tangibility, this is frustrating for the listener. Our experience of higher-order ambisonics leads us to suppose that this problem is soluble. In spite of the above, we strongly suspect that this kind of crude '2.5 dimensional' (cells are stacked horizontally but not vertically) analysis and synthesis method does offer a possible upward scalability. That is, this first crude attempt could be extended to many Soundfield microphones and representation cells. Further, synthetic material (or 'spot-miked material) can be injected into the scene.

We have also experimented with non-conventional centre-less shapes of sound field including a section of a street scene, a bowling alley, a waterfall and stream and a section of a Formula three racetrack. The objective in each was to produce a scene that can satisfy the ambulant, exploring perceiver so that there is no ideal listening position. These examples required a deal of synthetic "spatial Foley", an approach quite unlike traditional sound field sampling techniques.

5. FUTURE WORK

Synthesising such large-scale environments, whilst technically challenging, is fast becoming feasible as much audio software is being designed to handle more signal channels controlled by powerful computers.

The design of encoding environments similar to that implemented in The Morning Line Audio authoring software [28], will provide particularly interesting challenges. Especially, the control of location, movement and techniques to define spatial environments in such a large environment requires fundamental rethinking of current production approaches. Current pan-controls place images on the circumference of a circle (or surface of a sphere in the periphonic case). Dynamic panning to produce the impression of movement actually produces images in orbit around the listener, as though constrained by a piece of string. This is perceptually unsatisfying as listeners' *intuitive physics* [29] would expect normal movement of objects to tend to be in straight lines unless specially constrained.

In addition, we would like to explore 'multi-scale spatiality', where very large environments can also feature very fine-grain detail; the wind blowing a leaf along the road, for instance, or a whispered conversation. We are currently exploring hybrid approaches featuring second-order ambisonic cells managed within a *scene description* paradigm along with very small cross-talk-cancelling binaural fields, so that fine detail can be depicted in small listening regions that are nested within very large fields.

Finally, given that we know that perceptual attributes are intrinsically separable to form novel combinations, what experiences *could* we have? Composers, using perceptible materials, investigate aesthetic properties of a medium. The exciting challenge for composers is to use spatial sound to invent perceptual experiences that can be understood even though they are not *like* something in real environments. The twenty five composers who have written new works for The Morning Line have begun to explore some possibilities, but it is clear that there are significant opportunities for new work in this area. Could one dissipate an object into spatially-separated components, then reassemble them into another object in another location? Can one have an auditory equivalent of a magnifying glass so that we could 'zoom in' on a scene? Is a spatial equivalent to Shephard-Risset tones available, with endless spatial descents or even endless approaching? These notions embody *intentionality* and this produces behaviour that is not best addressed at the level of physical attributes.

To attempt some of these constituents may require a causality engine with artificial ontological foundations, yet conceptually, the leap likely to be small in comparison to what Plato achieved.

6. CONCLUSIONS

Plato's two-and-a-half thousand-year-old thought experiment remains germane to the problems of artificial environments. Artificial environments cannot be as ontologically rich as real environments –especially as we still do not know the ontological constituents of real environments. Perception, meanwhile, remains the ultimate "black box" problem, with immeasurably complex inputs, processes and outputs. Nevertheless, artificial environments can make available to us experiences and perceptual abstractions that would not otherwise be available.

Our proposals towards the pragmatic deployment of existing sound reproduction technologies, informed by perceptually significant spatial cartoons, sound images approaching their original scale and incorporating sound field attributes which allow listeners to explore their audio environment, have been implemented in the projects described above and appear to support significantly improved spatial audio perception.

7. REFERENCES

[1] Plato (360 BC) The Republic (Book VII) trans. Benjamin Jowett., available online, retrieved Sept 2004 from http://classics.mit.edu/Plato/republic.html

[2] Gaver W.W. (1993a) "What in the world do we hear? An ecological approach to auditory source perception" in: *Ecological Psychology 5,* pp. 1-29.

[3] Gaver W.W. (1993b) "How do we hear in the world? Explorations in ecological acoustics" in *Ecological Psychology* 5, pp.285-313.

[4] Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.

[5] Kahneman, D. Tversky, F., Krantz, (2002) "Maps of Bounded Rationality" Nobel Prize Lecture)

[6] Demonstration of walkers' gait derived via motion capture techniques. See: http://www.biomotionlab.ca/Demos/BMLwalker.html

[7] Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, 2:371-387

[8] Troje, N. F., Westhoff, C., and Lavrov, M.(2005) Person identification from biological motion: Effects of structural and kinematic cues. *Perception & Psychophysics* 67:667-675

[9] Rauschecker, J. P. and Tian, B., (2000) "Mechanisms and streams for processing of "what" and "where" in auditory cortex" Colloquium Paper in *PNAS.*, vol. 97 no. 22 11800-11806

[10] Eilan, N., McCarthy, R.A. & Brewer, W. (1993) *Spatial Representation: Problems in Philosophy and Psychology*. Oxford. Blackwell pp23-96

[11] Simons, D.J., & Levin, D.T. (1997). "Change Blindness". In: *Trends in Cognitive Science*, 1, pp 261–267.

[12] Neisser U, 1979 ``The control of information pickup in selective looking'', in *Perception and its Development: A Tribute to Eleanor J Gibson* Ed. A D Pick Hillsdale, NJ: Lawrence Erlbaum Associates pp 201 – 219

[13] Mack A, and Rock I, (1998) *Inattentional Blindness* Cambridge, MA. MIT Press

[14] Simons, D. & Chabris, C., (1999). "Gorillas in our midst: sustained inattentional blindness for dynamic events". *Perception*, 28, 1059-1074

[15] Strayer, D. L., Drews, F. A., Albert, R. W., & Johnston, W. A. (2001). "Cell phone induced perceptual impairments during simulated driving". In D. V. McGehee, J. D. Lee, & M. Rizzo (Eds.) *Driving Assessment 2001: International Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design.*

[16] Shinn-Cunningham, B. (2000) "Learning Reverberation: Considerations for Spatial Auditory Displays" in *Proc.6th. International Conference on Auditory Display*

[17] Glasgal, R.(20030 Surround Ambiophonic Recording and Reproduction in *Proceedings of the 24th international conference of the Audio Engineering Society,* Banff, Canada

[18] Farina, A (1993) "An example of adding spatial impression to recorded music: signal convolution with binaural impulse responses" in *Proc. of International Conference "Acoustics and recovery of spaces for music"* Ferrara, Italy.

[19] Pulkki 2006 "Directional audio coding in spatial sound reproduction and stereo upmixing," in *Proceedings of The AES 28th International Conference*, pp. 251-258, Piteå, Sweden

[20] Jot, J.-M., and Walsh, M. (2006). Binaural simulation of complex acoustic scenes for interactive audio. *In 121th AES Convention, San Francisco, USA. Preprint 6950*.

[21] Faller, C. (2006) Parametric Multichannel Audio Coding: Synthesis of Coherence Cues, *IEEE Transactions on Speech and Audio Processing*, Vol. 14, Nr. 1, pp. 299-310, 2006. .

[22] Glade Music Electronic Dance Festival 2006, UK

[23] Leonard, J. "When Geese Go Bad" Ambisonic recording. http://www.ambisonia.com/Members/soundmanjohn .

[24] Jacob Hoffman (2001) Concert at the WORM, MAXIS-Festival 2 (Music Alternative X-disciplinary approaches In Sound) at the Hallam University, Sheffield UK

[25] Martens, W. L. (2001) Two-subwoofer reproduction enables increased variation in auditory spatial imagery, *Proceedings of the Second International Workshop on Spatial Media* (IWSM01), University of Aizu, Japan.

[26] Myatt A and Malham DG (1995): Three Dimensional Sound Spatialisation Using Ambisonic Techniques, *Computer Music Journal,* Vol.19 No.4, MIT Press, Cambridge MA,USA, 1995, pp58-70

[27] Pulkki, V. (2002) "Compensating displacement of amplitude-panned virtual sources." Audio Engineering Society 22th Int. Conf. on Virtual, Synthetic and Entertainment Audio pp. 186-195. 2002 Espoo, Finland.

[28] Malham, D. G. Myatt, A. Larkin, O.. Worth, P. Paradis, Ma. (2010) "Audio Spatialization for The Morning Line"", Proc of the 128th Audio Engineering Society Convention, May, London, UK..

[29] Campbell, J. (1993) "The role of physical objects in spatial thinking" in: *Spatial Representation* Eds., Eilan, N., McCarthy, R., and Brewer, B. Oxford. Oxford University Press